Automation Bias in Intelligent Time Critical Decision Support Systems

M.L. Cummings[*]
Massachusetts Institute of Technology, Cambridge, MA 02319

Various levels of automation can be introduced by intelligent decision support systems, from fully automated, where the operator is completely left out of the decision process, to minimal levels of automation, where the automation only makes recommendations and the operator has the final say. For rigid tasks that require no flexibility in decision-making and with a low probability of system failure, higher levels of automation often provide the best solution. However, in time critical environments with many external and changing constraints such as air traffic control and military command and control operations, higher levels of automation are not advisable because of the risks and the complexity of both the system and the inability of the automated decision aid to be perfectly reliable. Human-in-the-loop designs, which employ automation for redundant, manual, and monotonous tasks and allow operators active participation, provide not only safety benefits, but also allow a human operator and a system to respond more flexibly to uncertain and unexpected events. However, there can be measurable costs to human performance when automation is used, such as loss of situational awareness, complacency, skill degradation, and automation bias. This paper will discuss the influence of automation bias in intelligent decision support systems, particularly those in aviation domains. Automation bias occurs in decision-making because humans have a tendency to disregard or not search for contradictory information in light of a computer-generated solution that is accepted as correct and can be exacerbated in time critical domains. Automated decision aids are designed to reduce human error but actually can cause new errors in the operation of a system if not designed with human cognitive limitations in mind.

I. Introduction

A primary design consideration in the development of intelligent decision support systems is how automation can best support human decision makers, and what level of automation should be introduced into a decision support system [1, 2]. Allocating roles and functions between the human and the computer is critical in defining efficient and effective system architectures, especially in the context of human supervisory control. Human supervisory control is the process by which a human operator intermittently interacts with a computer, receiving feedback from and providing commands to a controlled process or task environment, which is connected to that computer. According to research examining human and machine capabilities in air traffic control, which is representative of human supervisory control domains in which decisions must be made under time-pressure with little room for error, humans and machines (we now call computers) possess the respective strengths listed in Table 1, known as Fitts List [3]. As systems grow more complex, the use of automation to help humans navigate complex

Table 1. Strengths of Humans and Computers in Decision Making

Humans are better at:	Computers are better at:
Perceiving patterns	Responding quickly to control tasks
Improvising and using flexible procedures	Repetitive and routine tasks
Recalling relevant facts at the appropriate time	Reasoning deductively
Reasoning inductively	Handling many complex tasks simultaneously
Exercising judgment	

[*] Boeing Assistant Professor, Aeronautics and Astronautics, 77 Massachusetts Ave Room 33-305, AIAA Member.

and large problem spaces like those in complex time critical human supervisory control domains will be needed. However, it is equally important to recognize the critical role that humans play in these tasks and allocate decision-making functions between humans and computers accordingly.

In general, the concept of levels of automation (Table 2) in human supervisory control refers to the balance of automation and human control in decision and action selection [4, 5]. These levels of automation (LOAs), highly relevant to decision support systems, range from fully automated, where the operator is completely left out of the decision process, to minimal levels of automation, where the automation only makes recommendations or only simply provides filtering mechanisms. For rigid tasks that require no flexibility in decision-making and with a low probability of system failure, higher levels of automation often provides the best solution [6]. A "black box" approach to full automation, in which the automation's decision making is completely transparent to the human, can be useful for redundant tasks that require no knowledge-based judgments such as autopilot systems. However, the subsequent lack of system understanding and loss of situational awareness that full automation can cause can lead to unanticipated effects for more complex tasks. Even partially automated systems can result in measurable costs in human performance, such as loss of situational awareness, complacency, skill degradation, and decision biases [4].

Table 2. Levels of Automation

Automation Level	Automation Description
1	The computer offers no assistance: human must take all decision and actions.
2	The computer offers a complete set of decision/action alternatives, or
3	narrows the selection down to a few, or
4	suggests one alternative, and
5	executes that suggestion if the human approves, or
6	allows the human a restricted time to veto before automatic execution, or
7	executes automatically, then necessarily informs humans, and
8	informs the human only if asked, or
9	informs the human only if it, the computer, decides to.
10	The computer decides everything and acts autonomously, ignoring the human.

I. Automation Bias

While humans are typically effective in naturalistic decision making scenarios in which they leverage experience to solve real world ill-structured problems under stress [7], they are prone to fallible heuristics and various decision biases that are heavily influenced by experience, framing of cues, and presentation of information [8]. For example, confirmation bias takes place when people seek out information to confirm a prior belief and discount information that does not support this belief [9]. Another decision bias, assimilation bias, occurs when a person who is presented with new information that contradicts a preexisting mental model, assimilates the new information to fit into that mental model [10]. Of particular concern in the design of intelligent decision support systems is the human tendency toward automation bias, which occurs when a human decision maker disregards or does not search for contradictory information in light of a computer-generated solution which is accepted as correct [1, 11]. Operators are likely to turn over decision processes to automation as much as possible due to a cognitive conservation phenomenon [12], and teams of people, as well as individuals, are susceptible to automation bias [13]. Human errors that result from automation bias can be further decomposed into errors of commission and omission. Automation bias errors of omission occur when humans fail to notice problems because the automation does not alert them, while errors of commission occur when humans erroneously follow automated directives or recommendations [11].

Many studies in aviation have demonstrated the human tendency toward automation bias, causing both errors of commission and omission. The following discussion will detail how automation bias has affected human performance in three broad categories of aviation-related intelligent decision support systems: computer-assisted route planning, critical event diagnosis and action, as well as time-sensitive resource allocation. As will be demonstrated, intelligent decision aids are designed to reduce both human workload and error but actually can cause new errors in the operation of a system.

A. Automation bias in computer-assisted route planning

Automation is a critical component of any planning task due to the computational power of computers that allows them to quickly solve resource allocation, scheduling, and route planning problems through complex optimization algorithms. The use of algorithms to generate "optimal" plans is widespread; however, computer-generated solutions are not always truly optimal and in some cases, not even correct. Known as "brittleness," automation decision support models in complex systems cannot account for all potential conditions or relevant factors which could result in erroneous or misleading suggestions [14, 15]. In addition, as problem spaces grow in complexity, it becomes more difficult for the human to not only understand whether or not a computer-generated solution is correct, but how any one variable, a combination of variables, or missing information influence the computer's solution. This inability of the human to understand complex algorithms only exacerbates the tendency towards automation bias.

For example, in a study examining commercial pilot interaction with automation in an enroute flight planning tool, pilots, when given a computer-generated plan, exhibited significant automation over-reliance causing them to accept flight plans that were significantly sub-optimal. When presented with an automated solution (LOA 4, Table 2), 40% of pilots reasoned less or none at all when confronted with uncertainty in the problem space and deferred to erroneous automation recommendations, even though they were provided with tools with which to explore the automation space. The authors of this study suggest that even if automated critiquing alerts are provided to warn against possible constraint violations and/or provide suggestions to avoid constraint violations, human decision makers can be susceptible to automation bias [16]. In a similar experiment looking at levels of automated assistance in a military in-flight replanning task, pilots with LOA 4 assistance exhibited overreliance in the form of complacency. In this study, an automated decision aid planned a route taking into consideration time to targets, possible threats, and fuel state and subsequently presented pilots with its "optimal" solution, which could always be significantly improved through human intervention. Despite having the ability to change and improve the computer's solutions, subjects tended to accept the computer's solution without question [17].

In another study evaluating an automated decision support tool designed to aid pilots in generating trajectories to runways during emergencies, pilots tended toward overreliance and automation bias with a computer recommendation (LOA 4) [18]. In this study, pilots flying under a simulated emergency condition were presented with three different planning capabilities: 1) Charts, approach plates, and a circular slide-rule flight computer, 2) An emergency flight planner (EFP) which provided the pilot with a visual representation of the pilot's planned trajectory, as well as recalculations of critical parameters such as fuel remaining, and 3) An EFP which provided a recommendation in the form of a feasible trajectory for landing, which then the pilot could accept, ignore, clear, or modify as desired. Thus the three planning scenarios ranged from LOA 1 to LOA 4. In one test scenario, the EFP made erroneous predictions that did not account for degraded flight controls, thus the EFP's recommendation was erroneous. As a result of the LOA 4 recommendation, pilots exhibited automation bias and tended to follow the EFP's trajectory prediction, even when it was incorrect, with corresponding negative impact on performance.

B. Automation bias in critical event diagnosis and action

The use of automation to provide assistance in the diagnosis of system problems, as well as recommendations for actions, can be found not only in aviation domains, but also in medical, process control, and any other domain that requires monitoring of system and sub-system operation. Just as in the route planning examples, automated recommendations for diagnosis and action can also lead to complacency and overreliance in human decision makers. For example, in a study comparing decision-making with and without an automated monitoring aid (AMA), the reliability of the AMA significantly influenced the human decision error rate. In this study simulating monitoring tasks of pilots flying en-route, the purpose of the AMA was to notify subjects when certain geographic points were passed, as well as alert the pilot to systems that were not operating in their normal ranges. When the AMA was reliable, subjects performed better with an AMA than without one. However, several preprogrammed instances of omission (failure to notify operators of automation degradation) caused much higher error rates in the automated versus the non-automated condition (41% vs. 3%). When errors of commission were intentionally introduced (incorrect recommendation for action, LOA 4) subject error rate increased to 65%, despite the presence of 100%

reliable secondary contraindications. Thus, when automated monitoring aids operated reliably, they led to improved human performance and fewer errors as opposed to not having an aid. However, when the automation failed to detect and notify operators of an event, or incorrectly recommended action despite the availability of reliable confirmatory evidence, human error rates increased [19].

In another study examining the effectiveness of computer-generated recommendations on pilots' ability to diagnose and counter in-flight icing problems, the results were similar to the AMA study. Pilots were presented with either status displays (which merely present information, LOA 2) that indicated where ice was building on aircraft surfaces, or command displays (which recommend action, LOA 4). The command display indicated to the pilot the correct power and flap settings, as well as the proper pitch attitude. When the computer provided accurate advice, pilots with the command display performed better than pilots with the status display. However, when the computer's advice was erroneous, those people without the command display outperformed those with one [20]. Because of these results, it was recommended that unless decision aids are perfectly reliable, status displays should be used instead of command displays.

C. Automation bias in time-sensitive resource allocation

As mentioned previously, computer algorithms are extensively used to optimize complex scheduling and resource allocation problems in domains such as shipping and air traffic management. Human interaction problems with path planning and automation are similar to those found in scheduling and resource allocation optimization problems, in that humans have difficulty understanding whether or not a solution is truly "optimal". A resource allocation optimization problem becomes more complex and susceptible to automation bias in time-pressure situations like those that typically occur in command and control scenarios. As an example, in the recent development of an interface designed for supervision and resource allocation of in-flight GPS guided Tomahawk missiles, a design decision was required as to whether LOA 3 or 4 would be used to provide decision support for a human operator redirecting missiles in real time [21]. Operators were required to determine which candidate missile out of a pool of 8-16 missiles would be the correct missile to redirect to a time critical emergent target such as a mobile surface-air missile site. Some level of automated decision support was needed as a number of constraints had to be taken into account to make a correct decision: such as time to new target, fuel remaining, rules of engagement, warhead requirements, etc. The design question was whether or not the computer should provide the operator with ranked recommendations, i.e. recommending the most "optimal" missile given the situation, which corresponds to LOA 4. The alternative design was the incorporation of LOA 3 in which the computer simply filtered out all missiles that were not candidate missiles based only on physical constraints such as fuel remaining, and then allow the operator to weigh the remaining options to come to a decision with no ranked recommendations. The hypothesis was that LOA 4 would provide faster decisions, an important consideration given the time critical nature of redirecting missiles in flight. However, given the human propensity for automation bias, it was also likely that operators would accept the computer recommendation without seeking any disconfirming evidence and possibly make erroneous decisions.

Using low fidelity paper prototypes, subjects were split into two groups; those that made decisions to redirect missiles with a ranked list of missiles to include a highest candidate recommendation, and those that were simply presented with a list of missiles that could physically reach the target and met basic requirements. For half of the subjects with the ranked recommendations, other information was available on the same interface that made the automated recommendation the incorrect choice. Thus, the independent variables were whether or not the subject had the computer's recommendation and if the recommendation was correct or not. Two dependent variables, decision time and decision accuracy, were measured.

Those subjects with the ranked recommendations made decisions significantly faster than those without the computer's choice, and there was no difference in answer accuracy when the automated recommendations were correct, i.e. as long as the automation was correct, both groups generated the same number of correct answers, but those with the recommendations came to their decisions faster. However, for the case in which other sensor information conflicted with the computer-generated recommendations, subjects' accuracy of answers was significantly different than if the recommendation was accurate. Subjects generally selected the correct response when presented with reliable computer-generated recommendations; however, when presented with a recommendation that conflicted with other instructions, the number of incorrect responses increased, demonstrating a tendency towards automation bias. These results mirror those of the studies discussed in the critical event diagnosis category. Because of the ambiguous preliminary LOA decision -aiding results and the established tendency toward automation bias, the subsequent high fidelity prototype did not include the ranked recommendations.

II. Conclusion

While these studies demonstrate clear evidence of automation bias in laboratory settings, there is unfortunately ample anecdotal evidence in the "real world" where the consequences were deadly. As an example of automation bias in critical event diagnosis, in 1972, an Eastern L-1011crashed into the Florida Everglades, most likely in part due to an automation bias omission error. Upon execution of the landing checklist, the nose gear indicated unsafe. After engaging the autopilot, the crew intently focused on the unsafe nose gear, failing to notice several minutes later a gradual descent in altitude, which was likely caused by one of the pilots inadvertently bumping the control stick and disengaging the autopilot. The crew mistakenly relied on the automation to both keep the plane at the correct altitude and to warn them if the autopilot failed, leading to deaths of 101 crew and passengers [22].

Automation bias in time sensitive resource allocation was seen recently in the 2004 war with Iraq when the U.S. Army's Patriot missile system engaged in fratricide, shooting down a British Tornado and an American F/A-18, killing three aircrew. The system was designed to operate under LOA 6 (Table 2), and operators were given ten seconds to veto a computer solution. Unfortunately the displays were confusing and often incorrect, and operators were admittedly lacking training in a highly complex system [23]. Given the laboratory evidence that given an unreliable system, humans are still likely to approve computer-generated recommendations, it is not surprising that under the added stress of combat, Patriot operators did not veto the computer's solution.

As these cases demonstrate, automation bias is a very real concern in the development of intelligent systems that provide decision support for humans, especially those in time critical domains. Designers of intelligent systems should be aware of the potential negative effects of increasing levels of automation and the removal of the human decision maker from the control loop. Understanding where automated processes could occur and what levels would be appropriate for particular tasks is a key first step in avoiding potential problematic human-computer interactions. In systems like those in air traffic management and military command and control domains that require decision-making in environments with many external and dynamic constraints, higher levels of automation are not advisable because of the inability of automated decision aids to be perfectly reliable and the propensity for biased decision making.

However, designing a successful intelligent decision support system that leverages both the strengths of automation and the human as detailed in Table 1 is not simply a function of automation level. As demonstrated by several of the previously discussed studies, system reliability is also a significant contributor to automation bias and the success of an intelligent decision support system. In a recent study, one intervention used to ameliorate the effects of automation bias was to display dynamic system reliability. Pilots engaged in an experiment similar to the icing study discussed in the critical event diagnosis and action section were given either an overall system reliability for an intelligent neural net-based decision aid, or they were presented with trends of system confidence for the same decision aid. Pilots with the reliability trend display exhibited less automation bias and fewer ice-induced stalls than those who were simply given the overall system reliability [24]. While preliminary, these results are promising but more research is needed in the setting of reliability thresholds and the impact of such displays on a user's sense of trust for a system, which could be negatively impacted.

Intelligent decision aids are intended to reduce human error and workload but designers must be mindful that higher levels of automation combined with unreliable systems can actually cause new errors in system operation if not designed with human cognitive limitations and biases in mind. Design of an intelligent system that provides decision support must consider the human not just as a peripheral device, but also as an integrated system component that in the end, will ultimately determine the success or the failure of the system itself.

References

[1] Parasuraman, R. and Riley, V., Humans and Automation: Use, Misuse, Disuse, Abuse, *Human Factors*, 39, 1997, 230-253.

[2] Billings, C. E., *Aviation Automation: The Search For A Human-Centered Approach*, edited, Lawrence Erlbaum Associates, Hillsdale, N.J, 1997.

[3] Chapanis, A., Frick, F. C., Garner, W. R., Gebhard, J. W., Grether, W. F., Henneman, R. H., Kappaif, W. E., Newman, E. B., and Williams, A. C., Human Engineering for an effective air navigation and traffic control system, P. M. Fitts Eds., National Research Council Washington DC, 1951

[4] Parasuraman, R., Sheridan, T. B., and Wickens, C. D., A Model for Types and Levels of Human Interaction with Automation, *IEEE Transactions on Systems, Man, and Cybernetics*, 30, 2000, 286-297.

[5] Sheridan, T. B. and Verplank, W., "Human and Computer Control of Undersea Teleoperators", Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT, Cambridge, MA, 1978.

[6] Endsley, M. R. and Kaber, D. B., Level of automation effects on performance, situation awareness and workload in a dynamic control task, *Ergonomics*, 42, 1999, 462-492.

[7] Zsambok, C. E., Beach, R. B., and Klein, G., "Literature Review of Analytical and Naturalistic Decision Making", Naval Command, Control and Ocean Surveillance Center, San Diego, 1992.

[8] Tversky, A. and Kahneman, D., Judgment under Uncertainty: Heuristics and Biases, *Science*, 185, 1974, 1124-1131.

[9] Lord, C. G., Ross, L., and Lepper, M., Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence, *The Journal of Personality and Social Psychology*, 47, 1979, 1231-1243.

[10] Carroll, J. M. and Rosson, M. B., Paradox of the Active User, in Interfacing Thought: Cognitive Aspects of Human-Computer Interaction, J. M. Carroll Eds., MIT Press Cambridge, MA, 1987, 80-111.

[11] Mosier, K. L. and Skitka, L. J., Human Decision Makers and Automated Decision Aids: Made for Each Other? , in Automation and Human Performance: Theory and Applications, M. Mouloua Eds., Lawrence Erlbaum Associates, Inc. Mahwah, New Jersey, 1996, 201-220.

[12] Fiske, S. T. and Taylor, S. E., *Social cognition*, edited, McGraw-Hill, New York, 1991.

[13] Skitka, L. J. and Mosier, K. L., Automation Bias and Errors: Are Crews Better Than Individuals? , *International Journal of Aviation Psychology*, 10, 2000, 85-97.

[14] Smith, P., McCoy, E., and C. Layton, Brittleness in the design of cooperative problem-solving systems: The effects on user performance, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27, 1997, 360-371.

[15] Guerlain, S., Smith, P., Obradovich, J., Rudmann, S., Strohm, P., Smith, J., and Svirbely, J., Dealing with brittleness in the design of expert systems for immunohematology, *Immunohematology*, 12, 1996, 101-107.

[16] Layton, C., Smith, P. J., and McCoy, E., Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation, *Human Factors*, 36, 1994, 94-119.

[17] Johnson, K., Ren, L., Kuchar, J., and Oman, C., Interaction of Automation and Time Pressure in a Route Replanning Task, *International Conference on Human-Computer Interaction in Aeronautics*, Cambridge, MA, 2002. 132-137.

[18] Chen, T. L. and Pritchett, A. R., Development and evaluation of a cockpit decision-aid for emergency trajectory generation, *Journal of Aircraft*, 38, 2001, 935 - 943.

[19] Skitka, L. J., Mosier, K. L., and Burdick, M. D., Does automation bias decision-making? , *International Journal of Human-Computer Studies*, 51, 1999, 991-1006.

[20] Sarter, N. B. and Schroeder, B., Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing, *Human Factors*, 43, 2001, 573-583.

[21] Cummings, M. L., 'Designing Decision Support Systems for Revolutionary Command and Control Domains', Ph.D. Dissertation, Systems and Information Engineering, University of Virginia, Charlottesville, 2003.

[22] NTSB, "Aircraft Accident Report, Eastern Airlines Inc, L-1011 N310EA, Miami FL, December 19, 1972", AAR-73-14, NTSB, Washington, DC, 1973.

[23] 32nd Army Air and Missile Defense Command, in 'Patriot Missile Defense Operations during Operation Iraqi Freedom', Washington DC, 2003.

[24] McGuirl, J. M. and Sarter, N. B., How are we doing?: Presenting System Confidence Information to Support Trust Calibration and Adaptive Function Allocation., *Human Factors and Ergonomics Society 47th Annual Meeting*, Denver, CO, 2003.